

THE LAW OF SERIES

T. DOWNAROWICZ AND Y. LACROIX

January 8, 2006

ABSTRACT. We prove a general ergodic-theoretic result concerning the return time statistic, which, properly understood, sheds some new light on the common sense phenomenon known as *the law of series*. Let $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ be an ergodic process on finitely many states, with positive entropy. We show that the distribution function of the normalized waiting time for the first visit to a small cylinder set B is, for majority of such cylinders and up to epsilon, dominated by the exponential distribution function $1 - e^{-t}$. This fact has the following interpretation: The occurrences of such a “rare event” B can deviate from purely random in only one direction – so that for any length of an “observation period” of time, the first occurrence of B “attracts” its further repetitions in this period.

NOTE

This paper resulted from studying asymptotic laws for return/hitting time statistics in stationary processes, a field in ergodic theory rapidly developing in the recent years (see e.g. [A-G], [C], [C-K], [D-M], [H-L-V], [L] and the reference therein). Our result significantly contributes to this area due to both its generality and strength of the assertion. After having completely written the proof, during a free-minded discussion, the authors have discovered an astonishing interpretation of the result, clear even in terms of the common sense understanding of random processes. The organization of the paper is aimed to emphasise this discovery. The consequences for the field of asymptotic laws are moved toward the end of the paper.

INTRODUCTION

The phenomenon known as *the law of series* appears in many aspects of everyday life, science and technology. In the common sense understanding it signifies a sudden increase of frequency of a rare event, seemingly violating the rules of probability. Let us quote from Jean Moisset ([Mo]):

This law can be defined as: the repetition of identical or analogous events, things or symbols in space or time; for example: the announcement of several similar accidents on the same day, a series of strange events experienced by someone on the same day which are either happy ones (a period of good luck) or unfavorable (disastrous) ones, or the repetition of unexpected similar events. For example, you are invited to dinner and you are served a roast beef and you note that you were served the same menu the day before at your uncle’s home and the day before that at your cousin’s home.

1991 *Mathematics Subject Classification.* 37A50, 37A35, 37A05, 60G10.

Key words and phrases. stationary random process, positive entropy, return time statistic, hitting time statistic, repelling, attracting, limit law, the law of series.

This paper was written during the first author’s visit at CPT/ISITV, supported by CNRS. The research of the first author is supported by DBN grant 1 P03A 021 29

Another proverb describing more or less the same (with regard to unwanted events) is *misfortune seldom comes alone*. Both expressions exist in many languages, proving that the phenomenon has been commonly noticed throughout the world. In this setting it has been accounted to the category of unexplained mystery, parapsychics, parapsychology, together with “malignancy of fate”, “Murphy laws”, etc. Many pseudoscientific experiments have been conducted to prove obvious violation of statistical laws, where such laws were usually identified with the statistics of a series of independent trials ([St], [Km]). Equally many texts have been devoted to explain the anomaly within the framework of an independent process (see e.g., [Mi]), or merely as the weakness of our memory, keen to notice unusual events as more frequent just because they are more distinctive.

The phenomenon is also known in more serious science. In modeling and statistics it is sometimes called “clustering of data”. It is experimentally observed in many real processes, such as traffic jams, telecommunication network overloads, power consumption peaks, demographic peaks, stock market fluctuations, etc., as periods of increased frequency of occurrences of certain rare events. These anomalies are usually explained in terms of physical dependence (periods of propitious conditions) and complicated algorithms are implemented in modeling these processes to simulate them.

But, to our knowledge, there was no logical construction proving, in full generality, that there exists a “natural” tendency of rare events to appear in series, and the result we will present in this paper, or any of its possible variants, remained until now unnoticed by the specialists. We prove an ergodic-theoretic theorem on stationary stochastic processes, in which a wide range of “rare events” is shown to behave either in a way which we call “unbiased” (i.e., as in an independent process), or else exactly as it is specified in *the law of series*, i.e., so that the first occurrence increases the chances of untimely repetitions. Roughly speaking, we prove that **rare events appear in series whenever the unbiased behavior is perturbed: there is no other choice**. Besides ergodicity (which is automatically satisfied if we observe a chosen at random single realization of any process) we make only one essential, but obviously necessary, assumption on the process: it must maintain a “touch of randomness”, i.e., the future must not be completely determined by the past, which is equivalent to assuming positive entropy. Without this assumption a rotation of a compact group is an immediate example where events never appear in series. Of course, not every interesting rare event in reality can be modeled by the type of set we describe (cylinder over a long block), and we do not claim that our theorem fully explains the common sense phenomenon, but it certainly sheds on it some new light.

In terms of ergodic theory, we define two elementary antagonistic properties of the return times called “attracting” and “repelling”, and we prove that they behave quite differently in processes of zero and of positive entropy: attracting can persist for arbitrarily long blocks in both cases, while repelling must decay (as the length of blocks grows to infinity) in positive entropy processes. Many properties are known to differentiate between positive and zero entropy, but most of them involve a passage via measure-theoretic isomorphism, i.e., change of a generator, or require some additional structure. Our “decay of repelling” holds in general and for any finite generator, or even partition, as long as it generates positive entropy.

It is impossible not to mention here the theorem of Ornstein and Weiss [O-W2] which relates the return times of long blocks to entropy. However, this theorem says nothing about attracting or repelling, because the limit appearing in the statement

is insensitive to the proportions between gap sizes. Nevertheless, this remarkable result is very useful and it will help also in our proof. Our theorem's proof is entirely contained within the classics of ergodic theory; it relies on basic facts on entropy for partitions and sigma-fields, some elements of the Ornstein theory (ϵ -independence), the Shannon-McMillan-Breiman Theorem, the Ornstein-Weiss Theorem on return times, the Ergodic Theorem, basics of probability and calculus. We do not invoke any specialized machinery of stochastic processes or statistics.

The authors would like to thank Dan Rudolph for a hint leading to the construction of Example 2 and, in effect, to the discovery of the attracting/repelling asymmetry. We also thank Jean-Paul Thouvenot for his interest in the subject, substantial help, and the challenge to find a purely combinatorial proof (which we save for the future).

RIGOROUS DEFINITION AND STATEMENT

We establish the notation necessary to formulate the main result. Let $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ be an ergodic process on finitely many symbols, i.e., $\#\mathcal{P} < \infty$, σ is the standard left shift map and μ is an ergodic shift-invariant probability measure on $\mathcal{P}^{\mathbb{Z}}$. Most of the time, we will identify finite blocks with their cylinder sets, i.e., we agree that $\mathcal{P}^n = \bigvee_{i=0}^{n-1} \sigma^{-i}(\mathcal{P})$. Depending on the context, a block $B \in \mathcal{P}^n$ is attached to some coordinates or it represents a “word” which may appear in different places along the \mathcal{P} -names. We will also use the probabilistic language of random variables. Then $\mu\{R \in A\}$ ($A \subset \mathbb{R}$) will abbreviate $\mu(\{x \in \mathcal{P}^{\mathbb{Z}} : R(x) \in A\})$. Recall, that if the random variable R is nonnegative and $F(t) = \mu\{R \leq t\}$ is its distribution function, then the expected value of R equals $\int_0^\infty 1 - F(t) dt$.

For a set B of positive measure let R_B and \overline{R}_B denote the random variables defined on B (with the conditional measure $\mu_B = \frac{\mu}{\mu(B)}$) as the absolute and normalized first return time to B , respectively, i.e.,

$$R_B(y) = \min\{i > 0, \sigma^i(y) \in B\}, \quad \overline{R}_B(y) = \mu(B)R_B(y).$$

Notice that, by the Kac Theorem ([Kc]), the expected value of R_B equals $\frac{1}{\mu(B)}$, hence that of \overline{R}_B is 1 (that is why we call it “normalized”). We also define

$$G_B(t) = \int_0^t 1 - F_B(s) ds.$$

(The interpretation of this function is discussed in the following section.) Clearly, $G_B(t) \leq \min\{t, 1\}$ and the equality holds when $F_B(t) = 1_{[1, \infty)}$, that is, when B occurs precisely with equal gaps, i.e., periodically; the gap size then equals $\frac{1}{\mu(B)}$.

The key notions of this work are defined below:

Definition 1. We say that the visits to B *repel* (resp. *attract*) each other with intensity ϵ from a distance $t > 0$, if

$$G_B(t) \geq 1 - e^{-t} + \epsilon \quad (\text{resp. if } G_B(t) \leq 1 - e^{-t} - \epsilon).$$

We abbreviate that B *repels* (*attracts*) with intensity ϵ if its visits repel (attract) each other with intensity ϵ from some distance t .

Obviously, occurrences of an event may simultaneously repel from one distance and attract from another. Notice, that the maximal intensity of repelling is e^{-1} achieved at $t = 1$ when B appears periodically. The intensity of attracting can be arbitrarily close to 1 (when B appears in enormous clusters separated by huge pauses; see the next section). The main result follows:

Theorem 1. *If $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ is ergodic and has positive entropy, then for every $\epsilon > 0$ the measure of the union of all n -blocks $B \in \mathcal{P}^n$ which repel with intensity ϵ , converges to zero as n grows to infinity.*

We also provide an example (Example 2) in which, for a substantial collection of lengths, the majority of cylinders display strong attracting. Moreover, the process of Example 2 is isomorphic to a Bernoulli process, which implies that a partition with such strong attracting properties can be found in any measure-preserving transformation of positive entropy (see the Remark 3). Let us mention that it is easy to find zero entropy examples with either persistent repelling (discrete spectrum) or attracting (see the Example 3), or even both at a time (see the Remark 4).

INTERPRETATION AND ITS LIMITS

Let us elaborate a bit on the meaning of attracting and repelling for an event B . Let V_B be the random variable defined on X as the *hitting time statistic*, i.e., the waiting time for the first visit in B (the defining formula is the same as for R_B , but this time it is regarded on X with the measure μ). Further, let $\bar{V}_B = \mu(B)V_B$, called, by analogy, *the normalized hitting time* (although the expected value of this variable need not be equal to 1). By ergodicity, V_B and \bar{V}_B are well defined. By an elementary consideration of the skyscraper above B , one easily verifies, that the distribution function \tilde{F}_B of \bar{V}_B satisfies the inequalities:

$$G_B(t) - \mu(B) \leq \tilde{F}_B(t) \leq G_B(t)$$

(see [H-L-V] for more details). Because we deal with long blocks (so that, by the Shannon-McMillan-Breiman Theorem, $\mu(B)$ is, with high probability, very small), for sake of the interpretation, we will simply assume that $\tilde{F}_B = G_B$. Thus, attracting and repelling can be considered properties of the hitting rather than return time statistic. In fact, if we replace G_B by \tilde{F}_B in the definition of attracting/repelling, the formulation of Theorem 1 remains exactly the same, because it admits tolerance up to a fixed ϵ .

It is easy to see that if $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ is an independent Bernoulli process, then, for any long block B , $F_B(t) \approx 1 - e^{-t}$ (and also $\tilde{F}_B(t) \approx 1 - e^{-t}$) with high uniform accuracy. We will call such behavior “unbiased” (neither attracting nor repelling), and attracting and repelling can be viewed as deviations from the unbiased pattern.

Fix some $t > 0$. Consider the random variable I counting the number of occurrences of B in the time period $[0, \frac{t}{\mu(B)}]$. The expected value of I equals $\mu(B) \lfloor \frac{t}{\mu(B)} \rfloor \approx t$ (up to the ignorable error $\mu(B)$). On the other hand, $\mu\{I > 0\} = \mu\{V_B \leq \frac{t}{\mu(B)}\} = \tilde{F}_B(t)$. Attracting from the distance t occurs when the last value is smaller than it would be (say, for the same cylinder B) in an independent process. Because the expected value of I in the independent process is maintained (and equals approximately t), the conditional expected value of I on the set $\{I > 0\}$ must be larger in $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ than in the independent process. This fact can be interpreted as follows: If we observe the process for time $\frac{t}{\mu(B)}$ (which is our “memory length” or “lifetime of the observer”) and we happen to see the event B during this time at least once, then the expected number of times we will observe the event B is larger than the analogous value in the independent process. The first occurrence

repelling <i>B</i> <u><i>B</i></u> <i>B</i> <i>B</i> <u><i>B</i></u> <i>B</i> <u><i>B</i></u> <u><i>B</i></u> <u><i>B</i></u> <i>B</i> <i>B</i>
unbiased <i>B</i> <u><i>B</i></u> <i>B</i> <i>B</i> <u><i>B</i></u> <i>B</i> <u><i>B</i></u> <u><i>B</i></u> <u><i>B</i></u> <i>B</i> <i>B</i>
attracting <i>B</i> <u><i>B</i></u> <u><i>B</i></u> <u><i>B</i></u> <i>B</i> <u><i>B</i></u> <u><i>B</i></u> <u><i>B</i></u> <i>B</i> <i>B</i>
strong attr. <i>BBB</i> <i>BB</i> <i>B</i> <i>BBB</i> <i>BB</i> <i>BB</i>

Repelling from the distance t means exactly the opposite: The first occurrence lowers the expected number of repetitions within the observation period, i.e., repels them. If we have a mixed behavior, our impression about whether the event attracts or repels its repetitions depends on the length of our “memory”. Attracting not assisted by repelling (or assisted by repelling of an ignorably small intensity) means that no matter what memory length we apply, either we see a nearly unbiased behavior or the first occurrence visibly attracts further repetitions. Our Theorem 1 asserts that if we observe longer and longer blocks B , repelling from any distance must decay in both measure and intensity (while attracting can persist), so that for majority of long blocks we will see the behavior as described above.

Another consequence of attracting not assisted by repelling (or assisted by repelling of a very small intensity) is an increased variance of the return time statistic (the variance may even cease to exist). Thus, again, the gaps between the occurrences of B are driven away from the expected value, toward the extremities 0 and ∞ , and hence, into the pattern of clusters separated by longer pauses. We skip the elementary estimations of the variance.

It must be reminded: we do not claim, for any class of processes, that occurrences of long blocks will actually deviate from unbiased. There are conditions, weaker than full independence, under which the distributions of the normalized return times of long blocks converge almost surely to the exponential law. It is so, for instance, in Markov processes (with finite memory). In fact, such convergence is implied by a sufficient rate of mixing ([A-G], [H-S-V]). Yet, such processes seem to be somewhat exceptional and we expect that attracting rules in majority of processes (see the Question 5 at the end of the paper). As we have already mentioned, at least that much is true, that in any dynamical system with positive entropy there exist partitions with strong attracting properties.

It is important not to be misled by an oversimplified approach. The “decay of repelling” in positive entropy processes appears to agree with the intuitive understanding of entropy as chaos: repelling is a “self-organizing” property; it leads to a more uniform, hence less chaotic, distribution of an event along a typical orbit. Thus one might expect that repelling with intensity ϵ revealed by a fraction ξ of all n -blocks contributes to lowering an upper estimate of the entropy by some percentage proportional to ξ and depending increasingly on ϵ . If this happens for infinitely many lengths n with the same parameters ξ and ϵ , the entropy should be driven to zero by a geometric progression. Surprisingly, it is not quite so, and the phenomenon has more subtle grounds. We will present an example which exhibits the incorrectness of such intuition (see the Example 1 and the preceding discussion). Also, it will become obvious from the proof, that there is no gradual reduction of the entropy. The entropy is “killed completely in one step”, that means, positive entropy and persistent repelling lead to a contradiction by examining the blocks of one sufficiently large length n ; we do not use any iterated procedure requiring repelling for infinitely many lengths.

NOTATION AND PRELIMINARY FACTS

We now establish further notation and preliminaries needed in the proof. If $\mathbb{A} \subset \mathbb{Z}$ then we will write $\mathcal{P}^{\mathbb{A}}$ to denote the partition or sigma-field $\bigvee_{i \in \mathbb{A}} \sigma^{-i}(\mathcal{P})$. We will abbreviate $\mathcal{P}^n = \mathcal{P}^{[0,n]}$, $\mathcal{P}^{-n} = \mathcal{P}^{[-n,-1]}$, $\mathcal{P}^- = \mathcal{P}^{(-\infty,-1]}$ (a “finite future”, a “finite past”, and the “full past” of the process).

We assume familiarity of the reader with the basics of entropy for finite partitions and sigma-fields in a standard probability space. Our notation is compatible with [P] and we refer the reader to this book, as well as [Sh] and [W], for background and proofs. In particular, we will be using the following:

- * The entropy of a partition equals $H(\mathcal{P}) = -\sum_{A \in \mathcal{P}} \mu(A) \log_2(\mu(A))$.
- * For two finite partitions \mathcal{P} and \mathcal{B} , the conditional entropy $H(\mathcal{P}|\mathcal{B})$ is equal to $\sum_{B \in \mathcal{B}} \mu(B) H_B(\mathcal{P})$, where H_B is the entropy evaluated for the conditional measure μ_B on B .
- * The same formula holds for conditional entropy given a sub-sigma-field \mathcal{C} , i.e.,

$$\sum_{B \in \mathcal{B}} \mu(B) H_B(\mathcal{P}|\mathcal{C}) = H(\mathcal{P}|\mathcal{B} \vee \mathcal{C}).$$

- * The entropy of the process is given by any one of the formulas below

$$h = H(\mathcal{P}|\mathcal{P}^-) = \frac{1}{r} H(\mathcal{P}^r|\mathcal{P}^-) = \lim_{r \rightarrow \infty} \frac{1}{r} H(\mathcal{P}^r).$$

We will exploit the notion of ϵ -independence for partitions and sigma-fields. The definition below is an adaptation from [Sh], where it concerns finite partitions only. See also [Sm] for treatment of countable partitions. Because “ ϵ ” is reserved for the intensity of repelling, we will speak about β -independence.

Definition 2. Fix $\beta > 0$. A partition \mathcal{P} is said to be β -independent of a sigma-field \mathcal{B} if for any \mathcal{B} -measurable countable partition \mathcal{B}' holds

$$\sum_{A \in \mathcal{P}, B \in \mathcal{B}'} |\mu(A \cap B) - \mu(A)\mu(B)| \leq \beta.$$

A process $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ is called a β -independent process if \mathcal{P} is β -independent of the past \mathcal{P}^- .

A partition \mathcal{P} is independent of another partition or a sigma-field \mathcal{B} if and only if $H(\mathcal{P}|\mathcal{B}) = H(\mathcal{P})$. The following approximate version of this fact holds (see [Sh, Lemma 7.3] for finite partitions, from which the case of a sigma-field is easily derived).

Fact 1. *A partition \mathcal{P} is β -independent of another partition or a sigma-field \mathcal{B} if $H(\mathcal{P}|\mathcal{B}) \geq H(\mathcal{P}) - \xi$, for ξ sufficiently small. \square*

In course of the proof, a certain lengthy condition will be in frequent use. Let us introduce an abbreviation:

Definition 3. Given a partition \mathcal{P} of a space with a probability measure μ and $\delta > 0$, we will say that a property $\Phi(A)$ holds for $A \in \mathcal{P}$ with μ -tolerance δ if

$$\mu\left(\bigcup\{A \in \mathcal{P} : \Phi(A)\}\right) \geq 1 - \alpha.$$

We shall also need an elementary estimate, whose proof is an easy exercise.

Fact 2. *For each $A \in \mathcal{P}$, $H(\mathcal{P}) \leq (1 - \mu(A)) \log_2(\#\mathcal{P}) + 1$. \square*

In addition to the random variables of absolute and normalized return times R_B and \bar{R}_B , we will also use the analogous notions of the k^{th} absolute return time

$$R_B^{(k)} = \min\{i : \#\{0 < j \leq i : \sigma^j(y) \in B\} = k\},$$

and of the normalized k^{th} return time $\bar{R}_B^{(k)} = \mu(B)R_B^{(k)}$ (both defined on B), with $F_B^{(k)}$ always denoting the distribution function of the latter. Clearly, the expected value of $\bar{R}_B^{(k)}$ equals k .

THE IDEA OF THE PROOF AND THE BASIC LEMMA

Before we pass to the formal proof of Theorem 1, we would like to have the reader oriented in the mainframe of the idea behind it. We intend to estimate (from above, by $1 - e^{-t} + \epsilon$) the function G_{BA} , for long blocks of the form $BA \in \mathcal{P}^{[-n, r]}$. The “positive” part A has a fixed length r , while we allow the “negative” part B to be arbitrarily long. There are two key ingredients leading to the estimation. The first one, contained in Lemma 3, is the observation that for a fixed typical $B \in \mathcal{P}^{-n}$, the part of the process induced on B (with the conditional measure μ_B) generated by the partition \mathcal{P}^r , is not only a β -independent process, but it is also β -independent of many returns times $R_B^{(k)}$ of the cylinder B (see the Figure 2).

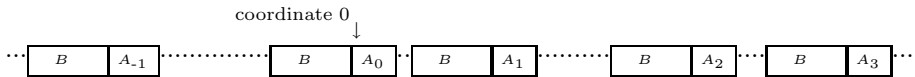


Figure 2: The process $\dots A_{-1}A_0A_1A_2\dots$ of r -blocks following the copies of B is a β -independent process with additional β -independence properties from the positioning of the copies of B .

This allows us to decompose (with high accuracy) the distribution function F_{BA} of the normalized return time of BA as follows:

$$\begin{aligned} F_{BA}(t) &= \mu_{BA}\{\bar{R}_{BA} \leq t\} = \mu_{BA}\{R_{BA} \leq \frac{t}{\mu(BA)}\} = \\ &= \sum_{k \geq 1} \mu_{BA}\{R_A^{(B)} = k, R_B^{(k)} \leq \frac{t}{p\mu(B)}\} \approx \sum_{k \geq 1} \mu_{BA}\{R_A^{(B)} = k\} \cdot \mu_B\{\bar{R}_B^{(k)} \leq \frac{t}{p}\} \approx \\ &= \sum_{k \geq 1} p(1-p)^{k-1} \cdot F_B^{(k)}\left(\frac{t}{p}\right), \end{aligned}$$

where $R_A^{(B)}$ denotes the first (absolute) return time of A in the process induced on B , and $p = \mu_B(A)$.

The second key observation is, assuming for simplicity full independence, that when trying to model some repelling for the blocks BA , we ascertain that it is largest, when the occurrences of B are purely periodic. Any deviation from periodicity of the B 's may only lead to increasing the intensity of attracting between the copies of BA , never that of repelling. We will explain this phenomenon more formally in a moment. Now, if B does appear periodically, then the normalized return time of BA is governed by the same geometric distribution as the normalized return time of A in the independent process induced on B . If p is small, this geometric distribution function becomes nearly the unbiased exponential law $1 - e^{-t}$. The smallness of p is *a priori* regulated by the choice of the parameter r (Lemma 1).

The phenomena that, assuming full independence, the repelling of BA is maximized by periodic occurrences of B , and that even then there is nearly no repelling, is captured by the following elementary lemma, which will be also useful later, near the end of the rigorous proof.

Lemma 0. *Fix some $p \in (0, 1)$. Let $F^{(k)}$ ($k \geq 1$) be a sequence of distribution functions on $[0, \infty)$ such that the expected value of the distribution associated to $F^{(k)}$ equals k . Define*

$$F(t) = \sum_{k \geq 1} p(1-p)^{k-1} F^{(k)}\left(\frac{t}{p}\right), \quad \text{and} \quad G(t) = \int_0^t 1 - F(s) ds.$$

Then $G(t) \leq \frac{1}{\log e_p} (1 - e_p^{-t})$, where $e_p = (1-p)^{-\frac{1}{p}}$.

Proof. We have

$$G(t) = \sum_{k \geq 1} p(1-p)^{k-1} \int_0^t 1 - F^{(k)}\left(\frac{s}{p}\right) ds.$$

We know that $F^{(k)}(t) \in [0, 1]$ and that $\int_0^\infty 1 - F^{(k)}(s) ds = k$ (the expected value). With such constraints, it is the indicator function $1_{[k, \infty)}$ that maximizes the integrals from 0 to t simultaneously for every t (because the “mass” k above the graph is for such choice of the function swept maximally to the left). The rest follows by direct calculations:

$$\begin{aligned} G(t) &\leq \sum_{k \geq 1} p(1-p)^{k-1} \int_0^t 1_{[0, k)}\left(\frac{s}{p}\right) ds = \int_0^t \sum_{k=\lceil \frac{s}{p} \rceil}^\infty p(1-p)^{k-1} ds = \\ &= \int_0^t (1-p)^{\lceil \frac{s}{p} \rceil} ds \leq \frac{(1-p)^{\frac{t}{p}} - 1}{\log(1-p)^{\frac{1}{p}}}. \quad \square \end{aligned}$$

Recall that the maximizing distribution functions $F_B^{(k)} = 1_{[k, \infty)}$ occur, for the normalized return time of a set B , precisely when B is visited periodically. This explains our former statement on this subject.

Let us comment a bit more on the first key ingredient, the β -independence. Establishing it is the most complicated part of the argument. The idea is to prove conditional (given a “finite past” \mathcal{P}^{-n}) β -independence of the “present” \mathcal{P}^r from jointly the full past and a large part of the future, responsible for the return times of majority of the blocks $B \in \mathcal{P}^{-n}$. But the future part must not be too large. Let us mention the existence of “bilaterally deterministic” processes with positive entropy (first discovered by Gurevič [G], see also [O-W1]), in which the sigma-fields generated by the coordinates $(-\infty, -m] \cup [m, \infty)$ do not decrease with m to the Pinsker factor; they are all equal to the entire sigma-field. (Coincidentally, our Example 1 has precisely this property; see the Remark 2.) Thus, in order to maintain any trace of independence of the “present” from our sigma-field already containing the entire past, its part in the future must be selected with an extreme care. Let us also remark that an attempt to save on the future sigma-fields by adjusting them individually to each block $B_0 \in \mathcal{P}^{-n}$ falls short, mainly because of the “off diagonal effect”; suppose \mathcal{P}^r is conditionally (given \mathcal{P}^{-n}) nearly independent of a sigma-field which determines the return times of only one selected block $B_0 \in \mathcal{P}^{-n}$. The independence still holds conditionally given any cylinder $B \in \mathcal{P}^{-n}$ from a collection of a large measure, but unfortunately, this collection can always miss the selected cylinder B_0 . In Lemmas 2 and 3, we succeed in finding a sigma-field (containing the full past and a part of the future), of which \mathcal{P}^r is conditionally β -independent, and which “nearly determines”, for majority of blocks $B \in \mathcal{P}^{-n}$, some finite number of their sequential return times (probably not all of them). This finite number is sufficient to allow the described earlier decomposition of the distribution function F_{BA} .

THE PROOF

Throughout the sequel we assume ergodicity and that the entropy h of $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ is positive. We begin our computations with an auxiliary lemma allowing us to assume (by replacing \mathcal{P} by some \mathcal{P}^r) that the elements of the “present” partition are small, relatively in most of $B \in \mathcal{P}^n$ and for every n . Note that the Shannon-McMillan-Breiman Theorem is insufficient: for the conditional measure the error term depends increasingly on n , which we do not fix.

Lemma 1. *For each δ there exists an $r \in \mathbb{N}$ such that for every $n \in \mathbb{N}$ the following holds for $B \in \mathcal{P}^{-n}$ with μ -tolerance δ :*

$$\text{for every } A \in \mathcal{P}^r, \quad \mu_B(A) \leq \delta.$$

Proof. Let α be so small that

$$\sqrt{\alpha} \leq \delta \quad \text{and} \quad \frac{h - 3\sqrt{\alpha}}{h + \alpha} \geq 1 - \frac{\delta}{2},$$

and set $\gamma = \frac{\alpha}{\log_2(\#\mathcal{P})}$. Let r be so big that

$$\frac{1}{r} \leq \alpha, \quad \frac{1}{r(h + \alpha)} \leq \frac{\delta}{2},$$

and that there exists a collection $\overline{\mathcal{P}^r}$ of no more than $2^{r(h+\alpha)} - 1$ elements of \mathcal{P}^r whose joint measure μ exceeds $1 - \gamma$ (by the Shannon-McMillan-Breiman Theorem).

Let $\widetilde{\mathcal{P}^r}$ denote the partition into the elements of $\overline{\mathcal{P}^r}$ and the complement of their union, and let \mathcal{R} be the partition into the remaining elements of \mathcal{P}^r and the complement of their union, so that $\mathcal{P}^r = \widetilde{\mathcal{P}^r} \vee \mathcal{R}$. For any n we have

$$\begin{aligned} rh &= H(\mathcal{P}^r | \mathcal{P}^-) \leq H(\mathcal{P}^r | \mathcal{P}^{-n}) = H(\widetilde{\mathcal{P}^r} \vee \mathcal{R} | \mathcal{P}^{-n}) = \\ &= H(\widetilde{\mathcal{P}^r} | \mathcal{R} \vee \mathcal{P}^{-n}) + H(\mathcal{R} | \mathcal{P}^{-n}) \leq H(\widetilde{\mathcal{P}^r} | \mathcal{P}^{-n}) + H(\mathcal{R}) \leq \\ &\quad \sum_{B \in \mathcal{P}^{-n}} \mu(B) H_B(\widetilde{\mathcal{P}^r}) + \gamma r \log_2(\#\mathcal{P}) + 1 \end{aligned}$$

(we have used Fact 2 for the last passage). After dividing by r , we obtain

$$\sum_{B \in \mathcal{P}^{-n}} \mu(B) \frac{1}{r} H_B(\widetilde{\mathcal{P}^r}) \geq h - \gamma \log_2(\#\mathcal{P}) - \frac{1}{r} \geq h - 2\alpha.$$

Because each term $\frac{1}{r} H_B(\widetilde{\mathcal{P}^r})$ is not larger than $\frac{1}{r} \log_2(\#\widetilde{\mathcal{P}^r})$ which was set to be at most $h + \alpha$, we deduce that

$$\frac{1}{r} H_B(\widetilde{\mathcal{P}^r}) \geq h - 3\sqrt{\alpha}$$

holds for $B \in \mathcal{P}^{-n}$ with μ -tolerance $\sqrt{\alpha}$, hence also with μ -tolerance δ . On the other hand, by Fact 2, for any B and $A \in \widetilde{\mathcal{P}^r}$, holds:

$$H_B(\widetilde{\mathcal{P}^r}) \leq (1 - \mu_B(A)) \log_2(\#\widetilde{\mathcal{P}^r}) + 1 \leq (1 - \mu_B(A))r(h + \alpha) + 1.$$

Combining the last two displayed inequalities we establish that, with μ -tolerance δ for $B \in \mathcal{P}^{-n}$ and then for every $A \in \widetilde{\mathcal{P}^r}$, holds

$$1 - \mu_B(A) \geq \frac{h - 3\sqrt{\alpha}}{h + \alpha} - \frac{1}{r(h + \alpha)} \geq 1 - \delta.$$

So, $\mu_B(A) \leq \delta$. Because \mathcal{P}^r refines $\widetilde{\mathcal{P}^r}$, the elements of \mathcal{P}^r are not larger. \square

We continue the proof with a lemma which can be deduced from [R1, Lemma 3]. We provide a direct proof. For $\alpha > 0$ and $M \in \mathbb{N}$ let

$$S(M, \alpha) = \bigcup_{m \in \mathbb{Z}} [mM + \alpha M, (m+1)M - \alpha M) \cap \mathbb{Z}.$$

Lemma 2. *For fixed α and r there exists M_0 such that for every $M \geq M_0$ holds,*

$$H(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) \geq rh - \alpha$$

(see the Figure 3).

*****○●.....*****.....*****.....*****.....

Figure 3. The circles indicate the coordinates 0 through $r - 1$, the conditioning sigma-field is over the coordinates marked by stars, which includes the entire past and part of the future with gaps of size $2\alpha M$ repeated periodically with period M (the first gap is half the size).

Proof. First assume that $r = 1$. Denote also

$$S'(M, \alpha) = \bigcup_{m \in \mathbb{Z}} [mM + \alpha M, (m+1)M) \cap \mathbb{Z}.$$

Let M be so large that $H(\mathcal{P}^{(1-\alpha)M}) < (1-\alpha)M(h+\gamma)$, where $\gamma = \frac{\alpha^2}{2(1-\alpha)}$. Then, for any $m \geq 1$,

$$H(\mathcal{P}^{S'(M, \alpha) \cap [0, mM)} | \mathcal{P}^-) \leq H(\mathcal{P}^{S'(M, \alpha) \cap [0, mM)}) < (1-\alpha)mM(h+\gamma).$$

Because $H(\mathcal{P}^{[0, mM)} | \mathcal{P}^-) = mMh$, the complementary part of entropy must exceed $mMh - (1-\alpha)mM(h+\gamma)$ (which equals $\alpha mM(h - \frac{\alpha}{2})$), i.e., we have

$$H(\mathcal{P}^{[0, mM) \setminus S'(M, \alpha)} | \mathcal{P}^- \vee \mathcal{P}^{S'(M, \alpha) \cap [0, mM)}) > \alpha mM(h - \frac{\alpha}{2}).$$

Breaking the last entropy term as a sum over $j \in [0, mM) \setminus S'(M, \alpha)$ of the conditional entropies of $\sigma^{-j}(\mathcal{P})$ given the sigma-field over all coordinates left of j and all coordinates from $S'(M, \alpha) \cap [0, mM)$ right of j , and because every such term is at most h , we deduce that more than half of these terms reach or exceed $h - \alpha$. So, a term not smaller than $h - \alpha$ occurs for a j within one of the gaps in the left half of $[0, mM)$. Shifting by j , we obtain $H(\mathcal{P} | \mathcal{P}^- \vee \sigma^i(\mathcal{P}^{S'(M, \alpha) \cap [0, \frac{mM}{2}]}) \geq h - \alpha$, where $i \in [0, \alpha M)$ denotes the relative position of j in the gap. As we increase m , one value i will repeat in this role along a subsequence m' . The operation \vee is continuous for increasing sequences of sigma-fields, hence $\mathcal{P}^- \vee \sigma^i(\mathcal{P}^{S'(M, \alpha) \cap [0, \frac{m'M}{2}]})$ converges over m' to $\mathcal{P}^- \vee \sigma^i(\mathcal{P}^{S'(M, \alpha)})$. The entropy is continuous for such passage, hence $H(\mathcal{P} | \mathcal{P}^- \vee \sigma^i(\mathcal{P}^{S'(M, \alpha)})) \geq h - \alpha$. The assertion now follows because $S(M, \alpha)$ is contained in $S'(M, \alpha)$ shifted to the left by any $i \in [0, \alpha M)$.

Finally, if $r > 1$, we can simply argue for \mathcal{P}^r replacing \mathcal{P} . This will impose that M_0 and M are divisible by r , but it is not hard to see that for large M the argument works without divisibility at a cost of a slight adjustment of α . \square

For a long block $B \in \mathcal{P}^{-n}$ let $((\mathcal{P}_B^r)^\mathbb{Z}, \mu_B, \sigma_B)$ denote the process induced on B generated by the restriction \mathcal{P}_B^r of \mathcal{P}^r to B (σ_B is the first return time map on B). The following lemma is the crucial item in our argument.

Lemma 3. *For every $\beta > 0$, $r \in \mathbb{N}$ and $K \in \mathbb{N}$ there exists n_0 such that for every $n \geq n_0$, with μ -tolerance β for $B \in \mathcal{P}^{-n}$, with respect to μ_B , \mathcal{P}^r is β -independent of jointly the past \mathcal{P}^- and the first K return times to B , $R_B^{(k)}$ ($k \in [1, K]$). In particular, $((\mathcal{P}_B^r)^\mathbb{Z}, \mu_B, \sigma_B)$ is a β -independent process.*

Proof. We choose ξ according to Fact 1, so that $\frac{\beta}{2}$ -independence is implied. Let α satisfy

$$0 < \frac{2\alpha}{h-\alpha} < 1, \quad 18K\sqrt{\alpha} < 1, \quad \sqrt{2\alpha} < \xi, \quad K\sqrt[4]{\alpha} < \frac{\beta}{2}.$$

Let n_0 be so large that $H(\mathcal{P}^r | \mathcal{P}^{-n}) < rh + \alpha$ for every $n \geq n_0$ and that for every $k \in [1, K]$ with μ -tolerance α for $B \in \mathcal{P}^{-n}$ holds

$$\mu_B\{2^{n(h-\alpha)} \leq R_B^{(k)} \leq 2^{n(h+\alpha)}\} > 1 - \alpha$$

(we are using Ornstein-Weiss Theorem [O-W2]; the multiplication by k is consumed by α in the exponent). Let $M_0 \geq 2^{n_0(h-\alpha)}$ be so large that the assertion of Lemma 2 holds for α , r and M_0 , and that for every $M \geq M_0$,

$$(M+1)^{1+\frac{2\alpha}{h-\alpha}} < \alpha M^2 \quad \text{and} \quad \frac{\log_2(M+1)}{M(h-\alpha)} < \alpha.$$

We can now redefine (enlarge) n_0 and M_0 so that $M_0 = \lfloor 2^{n_0(h-\alpha)} \rfloor$. Similarly, for each $n \geq n_0$ we set $M_n = \lfloor 2^{n(h-\alpha)} \rfloor$. Observe, that the interval where the first K returns of most n -blocks B may occur (up to probability α), is contained in $[M_n, \alpha M_n^2]$ (because $2^{n(h+\alpha)} \leq (M_n + 1)^{1+\frac{2\alpha}{h-\alpha}} < \alpha M_n^2$).

At this point we fix some $n \geq n_0$. The idea is to carefully select an M between M_n and $2M_n$ (hence not smaller than M_0), such that the initial K returns of nearly every n -block happen most likely inside (with all its n symbols) the set $S(M, \alpha)$, so that they are “controlled” by the sigma-field $\mathcal{P}^{S(M, \alpha)}$. Let $\alpha' = \alpha + \frac{n}{M_n}$, so that every n -block overlapping with $S(M, \alpha')$ is completely covered by $S(M, \alpha)$. By the second assumption on $M \geq M_0$ and by the formula connecting M_n and n , we have $\alpha' < 2\alpha$. To define M we will invoke the triple Fubini Theorem. Fix $k \in [1, K]$ and consider the probability space

$$\mathcal{P}^{-n} \times [M_n, 2M_n] \times \mathbb{N}$$

equipped with the (discrete) measure \mathcal{M} whose marginal on $\mathcal{P}^{-n} \times [M_n, 2M_n]$ is the product of μ (more precisely, of its projection onto \mathcal{P}^{-n}) with the uniform distribution on the integers in $[M_n, 2M_n]$, while, for fixed B and M , the measure on the corresponding \mathbb{N} -section is the distribution of the random variable $R_B^{(k)}$. In this space let S be the set whose \mathbb{N} -section for a fixed M (and any fixed B) is the set $S(M, \alpha')$. We claim that for every $l \in [M_n, \alpha M_n^2] \cap \mathbb{N}$ (and any fixed B) the $[M_n, 2M_n]$ -section of S has measure exceeding $1 - 16\alpha$. This is quite obvious (even for every $l \in [M_n, \infty)$ and with $1 - 15\alpha$) if $[M_n, 2M_n]$ is equipped with the normalized Lebesgue measure (see the Figure 4).

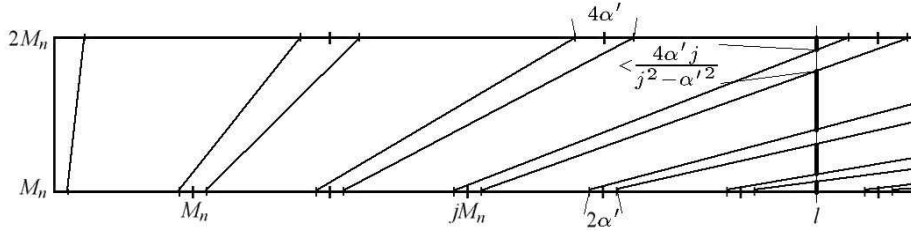


Figure 4: The complement of S splits into thin skew strips shown in the picture. The normalized Lebesgue measure of any vertical section of the j^{th} strip (starting at jM_n with $j \geq 1$) is at most $\frac{4\alpha'j}{j^2 - \alpha'^2} \leq \frac{5\alpha'}{j} \leq \frac{10\alpha}{j}$. Each vertical line at $l \geq M_n$ intersects strips with indices $j, j+1, j+2$ up to at most $2j$ (for some j), so the joint measure of the complement of the section of S does not exceed 15α .

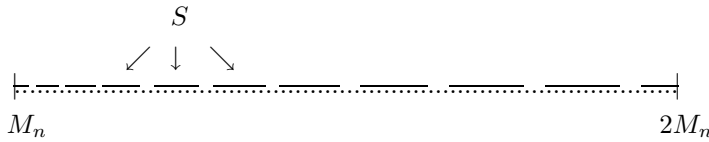


Figure 5: The discretization replaces the Lebesgue measure by the uniform measure on M_n integers, thus the measure of any interval can deviate from its Lebesgue measure by at most $\frac{1}{M_n}$. For $l \leq \alpha M_n^2$ the corresponding section of S (in this picture drawn horizontally) consists of at most αM_n intervals, so its measure can deviate by no more than α .

In the discrete case, however, *a priori* it might happen that the integers along some $[M_n, 2M_n]$ -section often “miss” the section of S leading to a decreased measure value. (For example, it is easy to see that for $l = (2M_n)!$ the measure of the section of S is zero.) But because we restrict to $l \leq \alpha M_n^2$, the discretization does not affect the measure of the section of S by more than α , and the estimate with $1 - 16\alpha$ holds (see the Figure 5 above).

Taking into account all other inaccuracies (the smaller than α part of S outside $[M_n, \alpha M_n^2]$ and the smaller than α part of S projecting onto blocks B which do not obey the Ornstein-Weiss return time estimate) it is safe to claim that

$$\mathcal{M}(S) > 1 - 18\alpha.$$

This implies that for every M from a set of measure at least $1 - 18\sqrt{\alpha}$ the measure of the $(\mathcal{P}^{-n} \times \mathbb{N})$ -section of S is larger than or equal to $1 - \sqrt{\alpha}$. For every such M , with μ -tolerance $\sqrt[4]{\alpha}$ for $B \in \mathcal{P}^{-n}$, the probability μ_B that the k^{th} repetition of B falls in $S(M, \alpha')$ (hence with all its n terms inside the set $S(M, \alpha)$) is at least $1 - \sqrt[4]{\alpha}$.

Because $18K\sqrt{\alpha} < 1$, there exists at least one M for which the above holds for every $k \in [1, K]$. This is our final choice of M . For this M , with μ -tolerance $K\sqrt[4]{\alpha}$, all considered K returns of B are, with probability $1 - \sqrt[4]{\alpha}$ (each), determined by the sigma-field $\mathcal{P}^{S(M, \alpha)}$. More precisely, for each $k \in [1, K]$ there is a set U_k of measure μ_B at most $\sqrt[4]{\alpha}$ such that the sets $\{R_B^{(k)} = i\} \setminus U_k$ agree with some $Q_i^{(k)} \setminus U_k$, where each $Q_i^{(k)}$ is $\mathcal{P}^{S(M, \alpha)}$ -measurable. Thus, we can modify the variable $R_B^{(k)}$ so it is $\mathcal{P}^{S(M, \alpha)}$ -measurable and equal to the original except on U_k . We denote such a modification by $\tilde{R}_B^{(k)}$.

Let us go back to our entropy estimates. We have, by Lemma 2,

$$\begin{aligned} \sum_{B \in \mathcal{P}^{-n}} \mu(B) H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) &= H(\mathcal{P}^r | \mathcal{P}^{-n} \vee \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) = \\ &H(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) \geq rh - \alpha \geq H(\mathcal{P}^r | \mathcal{P}^{-n}) - 2\alpha = \\ &\sum_{B \in \mathcal{P}^{-n}} \mu(B) H_B(\mathcal{P}^r) - 2\alpha. \end{aligned}$$

Because $H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) \leq H_B(\mathcal{P}^r)$ for every B , we deduce that with μ -tolerance $\sqrt{2\alpha}$ for $B \in \mathcal{P}^{-n}$ must hold

$$H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) \geq H_B(\mathcal{P}^r) - \sqrt{2\alpha} \geq H_B(\mathcal{P}^r) - \xi.$$

Combining this with the preceding arguments, with μ -tolerance $K\sqrt[4]{\alpha} + \sqrt{2\alpha} < \beta$ for $B \in \mathcal{P}^{-n}$ both the above entropy inequality holds, and we have the $\mathcal{P}^{S(M, \alpha)}$ -measurable modifications $\tilde{R}_B^{(k)}$ of the return times. By the choice of ξ , we obtain that with respect to μ_B , \mathcal{P}^r is jointly $\frac{\beta}{2}$ -independent of the past and the modified return times $\tilde{R}_B^{(k)}$ ($k \in [1, K]$). Because $\mu(\bigcup_{k \in [1, K]} U_k) \leq K\sqrt[4]{\alpha} < \frac{\beta}{2}$, this clearly implies β -independence if each $\tilde{R}_B^{(k)}$ is replaced by $R_B^{(k)}$. \square

To complete the proof of Theorem 1 it now remains to put the items together.

Proof of Theorem 1. Fix an $\epsilon > 0$. On $[0, \infty)$, the functions

$$g_p(t) = \min\{1, \frac{1}{\log e_p}(1 - e_p^{-t}) + pt\},$$

where $e_p = (1-p)^{-\frac{1}{p}}$, decrease uniformly to $1 - e^{-t}$ as $p \rightarrow 0^+$. So, let δ be such that $g_\delta(t) \leq 1 - e^{-t} + \epsilon$ for every t . We also assume that

$$(1 - 2\delta)(1 - \delta) \geq 1 - \epsilon.$$

Let r be specified by Lemma 1, so that $\mu_B(A) \leq \delta$ for every $n \geq 1$, every $A \in \mathcal{P}^r$ and for $B \in \mathcal{P}^{-n}$ with μ -tolerance δ . On the other hand, once r is fixed, the partition \mathcal{P}^r has at most $(\#\mathcal{P})^r$ elements, so with μ_B -tolerance δ for $A \in \mathcal{P}^r$, $\mu_B(A) \geq \delta(\#\mathcal{P})^{-r}$. Let \mathcal{A}_B be the subfamily of \mathcal{P}^r (depending on B) where this inequality holds. Let K be so large that for any $p \geq \delta(\#\mathcal{P})^{-r}$,

$$\sum_{k=K+1}^{\infty} p(1-p)^k < \frac{\delta}{2},$$

and choose $\beta < \delta$ so small that

$$(K^2 + K + 1)\beta < \frac{\delta}{2}.$$

The application of Lemma 3 now provides an n_0 such that for any $n \geq n_0$, with μ -tolerance β for $B \in \mathcal{P}^{-n}$, the process induced on B generated by \mathcal{P}^r has the desired β -independence properties involving the initial K return times of B . So, with tolerance $\delta + \beta < 2\delta$ we have both, the above β -independence and the estimate $\mu_B(A) < \delta$ for every $A \in \mathcal{P}^r$. Let \mathcal{B}_n be the subfamily of \mathcal{P}^{-n} where these two conditions hold. Fix some $n \geq n_0$.

Let us consider a cylinder set $B \cap A \in \mathcal{P}^{[-n, r]}$ (or, equivalently, the block BA), where $B \in \mathcal{B}_n$, $A \in \mathcal{A}_B$. The length of BA is $n + r$, which represents an arbitrary integer larger than $n_0 + r$. Notice that the family of such sets BA covers more than $(1 - 2\delta)(1 - \delta) \geq 1 - \epsilon$ of the space.

We will examine the distribution of the normalized first return time for BA . In addition to our customary notations of return times, let $R_A^{(B)}$ be the first (absolute) return time of A in $((\mathcal{P}_B^r)^\mathbb{Z}, \mu_B, \sigma_B)$, i.e., the variable defined on BA , counting the number of visits to B until the first return to BA . Let $p = \mu_B(A)$. We have

$$\begin{aligned} F_{BA}(t) &= \mu_{BA}\{\overline{R}_{BA} \leq t\} = \mu_{BA}\{R_{BA} \leq \frac{t}{\mu(BA)}\} = \\ &= \sum_{k \geq 1} \mu_{BA}\{R_A^{(B)} = k, R_B^{(k)} \leq \frac{t}{p\mu(B)}\}. \end{aligned}$$

The k^{th} term of this sum equals

$$\frac{1}{p} \mu_B(\{A_k = A\} \cap \{A_{k-1} \neq A\} \cap \cdots \cap \{A_1 \neq A\} \cap \{A_0 = A\} \cap \{R_B^{(k)} \leq \frac{t}{p\mu(B)}\}),$$

where A_i is the r -block following the i^{th} copy of B (the counting starts from 0 at the copy of B positioned at $[-n, -1]$).

By Lemma 3, for $k \leq K$, in this intersection of sets each term is β -independent of the intersection right from it. So, proceeding from the left, we can replace the probabilities of the intersections by products of probabilities, allowing an error of β . Note that the last term equals $\mu_B\{\overline{R}_B^{(k)} \leq \frac{t}{p}\} = F_B^{(k)}(\frac{t}{p})$. Jointly, the inaccuracy will not exceed $(K + 1)\beta$:

$$\left| \mu_{BA}\{R_A^{(B)} = k, R_B^{(k)} \leq \frac{t}{p\mu(B)}\} - p(1-p)^{k-1} F_B^{(k)}(\frac{t}{p}) \right| \leq (K + 1)\beta.$$

Similarly, we also have $\left| \mu_{BA}\{R_A^{(B)} = k\} - p(1-p)^{k-1} \right| \leq K\beta$, hence the tail of the series $\mu_{BA}\{R_A^{(B)} = k\}$ above K is smaller than $K^2\beta$ plus the tail of the geometric series $p(1-p)^{k-1}$, which, by the fact that $p \geq \delta(\#\mathcal{P})^{-r}$, is smaller than $\frac{\delta}{2}$. Therefore

$$F_{BA}(t) \approx \sum_{k \geq 1} p(1-p)^{k-1} F_B^{(k)}\left(\frac{t}{p}\right),$$

up to $(K^2 + K + 1)\beta + \frac{\delta}{2} \leq \delta$, uniformly for every t . By the application of Lemma 0, G_{BA} satisfies

$$G_{BA}(t) \leq \min\left\{1, \frac{1}{\log e_p}(1 - e_p^{-t}) + \delta t\right\} \leq g_\delta(t) \leq 1 - e^t + \epsilon$$

(because $p \leq \delta$). We have proved that for our choice of ϵ and an arbitrary length $m \geq n_0 + r$, with μ -tolerance ϵ for the cylinders $C \in \mathcal{P}^m$, the intensity of repelling between visits to C is at most ϵ . This concludes the proof of Theorem 1. \square

CONSEQUENCES FOR LIMIT LAWS

The studies of limit laws for return/hitting time statistics are based on the following approach: For $x \in \mathcal{P}^{\mathbb{Z}}$ define $F_{x,n} = F_B$ (and $\tilde{F}_{x,n} = \tilde{F}_B$), where B is the block $x[0, n)$ (or the cylinder in \mathcal{P}^n containing x). Because for nondecreasing functions $F : [0, \infty) \rightarrow [0, 1]$, the weak convergence coincides with the convergence at continuity points, and it makes the space of such functions metric and compact, for every x there exists a well defined collection of limit distributions for $F_{x,n}$ (and for $\tilde{F}_{x,n}$) as $n \rightarrow \infty$. They are called *limit laws for the return (hitting) times at x* . Due to the integral relation ($\tilde{F}_B \approx G_B$) a sequence of return time distributions converges weakly if and only if the corresponding hitting time distributions converge pointwise (see [H-L-V]), so the limit laws for the return times completely determine those for hitting times and *vice versa*. A limit law is *essential* if it appears along some subsequence (n_k) for x 's in a set of positive measure. In particular, the strongest situation occurs when there exists an almost sure limit law along the full sequence (n) . Most of the results concerning the limit laws, obtained so far, can be classified in three major groups: a) characterizations of possible essential limit laws for specific zero entropy processes (e.g. [D-M], [C-K]; these limit laws are usually atomic for return times or piecewise linear for hitting times), b) finding classes of processes with an almost sure exponential limit law along (n) (e.g. [A-G], [H-S-V]), and c) results concerning not essential limit laws, limit laws along sets other than cylinders (see [L]; every probabilistic distribution with expected value not exceeding 1 can occur in any process as the limit law for such general return times), or other very specific topics. As a consequence of our Theorem 1, we obtain, for the first time, a serious bound on the possible essential limit laws for the hitting time statistics along cylinders in the general class of ergodic positive entropy processes. The statement (1) below is even slightly stronger, because we require, for a subsequence, convergence on a positive measure set, but not necessarily to a common limit.

Theorem 2. *Assume ergodicity and positive entropy of the process $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$.*

- (1) *If a subsequence (n_k) is such that \tilde{F}_{x,n_k} converge pointwise to some limit laws \tilde{F}_x on a positive measure set A of points x , then almost surely on A , $\tilde{F}_x(t) \leq 1 - e^{-t}$ at each $t \geq 0$.*
- (2) *If (n_k) grows sufficiently fast, then there is a full measure set, such that for every x in this set holds: $\limsup_k \tilde{F}_{x,n_k}(t) \leq 1 - e^{-t}$ at each $t \geq 0$.*

Proof. The implication from Theorem 1 to Theorem 2 is obvious and we leave it to the reader. For (2) we hint that (n_k) must grow fast enough to ensure summability of the measures of the sets where the intensity of repelling persists. \square

EXAMPLES

The first construction will show that for each $\delta > 0$ and $n \in \mathbb{N}$ there exists $N \in \mathbb{N}$ and an ergodic process on N symbols with entropy $\log_2 N - \delta$, such that the n -blocks from a collection of joint measure equal to $\frac{1}{n}$ repel with nearly the maximal possible intensity e^{-1} . Because δ can be extremely small compared to $\frac{1}{n}$, this construction illustrates, that there is no “reduction of entropy” by an amount proportional to the fraction of blocks which reveal strong repelling.

Example 1. Let \mathcal{P} be an alphabet of a large cardinality N . Divide \mathcal{P} into two disjoint subsets, one, denoted \mathcal{P}_0 , of cardinality $N_0 = N2^{-\delta}$ and the relatively small (but still very large) rest which we denote by $\{1, 2, \dots, r\}$ (we will refer to these symbols as “markers”). For $i = 1, 2, \dots, r$, let \mathcal{B}_i be the collection of all n -blocks whose first $n-1$ symbols belong to \mathcal{P}_0 and the terminal symbol is the marker i . The cardinality of \mathcal{B}_i is N_0^{n-1} . Let \mathcal{C}_i be the collection of all blocks of length nN_0^{n-1} obtained as concatenations of blocks from \mathcal{B}_i using each of them exactly once. The cardinality of \mathcal{C}_i is $(N_0^{n-1})!$. Let X be the subshift whose points are infinite concatenations of blocks from $\bigcup_{i=1}^r \mathcal{C}_i$, in which every block belonging to \mathcal{C}_i is followed by a block from \mathcal{C}_{i+1} ($1 \leq i < r$) and every block belonging to \mathcal{C}_r is followed by a block from \mathcal{C}_1 . Let μ be the shift-invariant measure of maximal entropy on X . It is immediate to see that the entropy of μ is $\frac{1}{nN_0^{n-1}} \log_2((N_0^{n-1})!)$, which, for large N , nearly equals $\log_2 N_0 = \log_2 N - \delta$. Finally observe that the measure of each $B \in \mathcal{B}_i$ equals $\frac{1}{nrN_0^{n-1}}$, the joint measure of $\bigcup_{i=1}^r \mathcal{B}_i$ is exactly $\frac{1}{n}$, and every block B from this family appears in any $x \in X$ with gaps ranging between $\frac{1-\frac{1}{n}}{\mu(B)}$ and $\frac{1+\frac{1}{n}}{\mu(B)}$, revealing strong repelling.

Remark 1. Viewing blocks of length nrN_0^{n-1} starting with a block from \mathcal{C}_1 as a new alphabet, and repeating the above construction inductively, we can produce an example (with the measure of maximal entropy on the intersection of systems created in consecutive steps) with entropy $\log_2 N - 2\delta$, in which the strong repelling will occur with probability $\frac{1}{n_k}$ for infinitely many lengths n_k .

Remark 2. The process described in the above remark is (somewhat coincidentally; it was not designed for that) bilaterally deterministic: for every $m \in \mathbb{N}$ the sigma-field $\mathcal{P}(-\infty, -m] \cup [m, \infty)$ equals the full (product) sigma-field. Indeed, suppose we see all entries of a point x except on the interval $(-m, m)$. In a typical point, this interval is contained between a pair of successive markers i for some level k of the inductive construction. Then, by examining this point’s entries far enough to the left and right we will see completely all but one blocks from the family \mathcal{B}_i which constitute the block $C \in \mathcal{C}_i$ covering the considered interval. Because every block from \mathcal{B}_i is used in C exactly once, by elimination, we will be able to determine the missing block and hence all symbols in $(-m, m)$.

The next construction shows that there exists a process isomorphic to a Bernoulli process with an almost sure limit law $\hat{F} \equiv 0$ for the normalized hitting times (strong attracting), achieved along a subsequence of upper density 1. In particular, this answers in the negative a question of Zaqueu Coelho ([C]), whether all processes isomorphic to Bernoulli processes have necessarily the exponential limit law for the

hitting (and return) times. The idea of the construction was suggested to us by D. Rudolph ([R2]), who attributes the method to Arthur Rothstein.

Example 2. We will build a decreasing sequence of subshifts of finite type (SFT's). In each we will regard the measure of maximal entropy. Begin with the full shift X_0 on a finite alphabet. Select r words W_1, W_2, \dots, W_r of some length l and create r SFT's: $X_0^{(1)}, X_0^{(2)}, \dots, X_0^{(r)}$, forbidding one of these words in each of them, respectively. Choose a length n so large that in the majority of blocks of this length in $X_0^{(i)}$ all words W_j except W_i will appear at least once. Now choose another length m , such that in the majority of blocks of this length every block C of length between n and n^2 will appear many times. Now define X_1 as the subshift whose each point is a concatenation of the form $\dots B_1 B_2 \dots B_r B_1 \dots$, where B_i is a block appearing in $X_0^{(i)}$ of length either m or $m+1$. Obviously, a typical block C of any length between n and n^2 appearing in X_1 comes from some $X_0^{(i)}$, hence contains all W_j 's except W_i , therefore in a typical $x \in X_1$, C will appear many times within each component B_i representing $X_0^{(i)}$, and then it will be absent for a long time, until the next representative of $X_0^{(i)}$. So, every such block will reveal strong attracting. It is not hard to see that X_1 is a mixing SFT and its d-bar distance from the full shift is small whenever the length l of the (few) forbidden words W_i is large. We can now repeat the construction starting with X_1 , and radically increasing all parameters. We can arrange that the d-bar distances are summable, so the limit system X (the intersection of the X_k 's), more precisely its measure of maximal entropy, is also a d-bar limit. Each mixing SFT is isomorphic to a Bernoulli process and this property passes via d-bar limits (see [O], [Sh]), hence X is also isomorphic to a Bernoulli process. This system has the almost sure limit law $\tilde{F} \equiv 0$ for hitting times (or $F \equiv 1$ for the return times) achieved along a sequence containing infinitely many intervals of the form $[n, n^2]$. Such sequence has upper density 1.

Remark 3. It is also possible to construct a process X_h as above with any preassigned entropy h . On the other hand, it is well known ([Si]), that every measure-preserving transformation with positive entropy h possesses a Bernoulli factor of the same entropy. By the Ornstein Theorem ([O]) this factor is isomorphic to X_h . The generator of X_h appears as a partition of the space on which the initial measure-preserving transformation is defined. This proves the universality of “the law of series”: **in every measure-preserving transformation there exists a partition generating the full entropy, which has the “strong repelling properties”** (i.e., almost sure limit law $\tilde{F} \equiv 0$ along a sequence of lengths of upper density 1).

Various zero entropy processes with persistent repelling or attracting are implicit in the existing literature. Extreme repelling (with intensity converging to e^{-1} as the length of blocks grows) occurs for example in odometers, or, more generally, in rank one systems ([C-K]). For completeness, we sketch two zero entropy processes with features of positive entropy: repelling, and the unbiased behavior.

Example 3. Take the product of the independent Bernoulli process on two symbols with an odometer (modeled by an adequate process, for example a regular Toeplitz subshift; see [D] for details on Toeplitz flows). Call this product process X_0 . The odometer factor provides, for each $k \in \mathbb{N}$, markers dividing each element into so-called k -blocks of equal lengths p_k . Each p_k is a multiple of p_{k-1} and each k -block is a concatenation of $(k-1)$ -blocks. Now we create a new process X_1 by “stuttering”: if $x_0 \in X_0$ is a concatenation $\dots ABCD \dots$ of 1-blocks, we create $x_1 \in X_1$ as

$\dots AABBBCCDD \dots$, with the number of repetitions $q_1 = 2$. In X_1 the lengths of the k -blocks for $k > 1$ have doubled. Repeating the stuttering for 2-blocks of X_1 with a number of repetitions $q_2 \geq 2$, we obtain a process X_2 . And so on. Because in each step we reduce the entropy by at least half, the limit process has entropy zero. If the q_k 's grow sufficiently fast, we obtain, like in the previous example, a system with strong attracting for a set of lengths of upper density 1. Consider a modification of this example where $q_k = 2$ for each k and each pair AA (also BB , etc.) is substituted by $A\bar{A}$ ($B\bar{B}$, etc.), where \bar{A} is the "mirror" of A , i.e., with the symbols 0 are replaced by 1 and *vice versa*. It is not very hard to compute, that such process (although has entropy zero), has the same limit law properties as the independent process: almost sure convergence along the full sequence (n) to the unbiased (exponential) limit law.

Remark 4. It is not hard to construct zero entropy processes with persistent mixed behavior. For example, applying the "stuttering technique" to an odometer one obtains a process in which a typical block B occurs in periodically repeated pairs: $BB \dots BB \dots BB \dots$, i.e., with the function $G_B \approx \min\{1, \frac{t}{2}\}$ (which reveals attracting with intensity $\frac{\log 2 - 1}{2}$ at $t_1 = \log 2$ and repelling with intensity e^{-2} at $t_2 = 2$). We skip the details.

QUESTIONS

Question 1. Is there a speed of the convergence to zero of the joint measure of the "bad" blocks in Theorem 1? More precisely, does there exist a positive function $s(n, \epsilon, \#\mathcal{P})$ converging to zero as n grows, such that if for some ϵ and infinitely many n 's, the joint measure of the n -blocks which repel with intensity ϵ exceeds $s(n, \epsilon, \#\mathcal{P})$, then the process has necessarily entropy zero? (By the Example 1, $\frac{1}{n}$ is not enough.)

Question 2. Can one strengthen the Theorem 2 as follows:

$$\limsup_{n \rightarrow \infty} \tilde{F}_{x,n} \leq 1 - e^{-t} \quad \mu\text{-almost everywhere?}$$

Question 3. In Lemma 3, can one obtain \mathcal{P}^r conditionally β -independent of jointly the past and *all* return times $R_B^{(k)}$ ($k \geq 1$) (for sufficiently large n , with μ -tolerance β for $B \in \mathcal{P}^{-n}$)? In other words, can the β -independent process $((\mathcal{P}_B^r)^{\mathbb{Z}}, \mu_B, \sigma_B)$ be obtained β -independent of the factor-process generated by the partition into B and its complement?

Question 4. (suggested by J-P. Thouvenot) Find a purely combinatorial proof of Theorem 1, by counting the quantity of very long strings (of length m) inside which a positive fraction (in measure) of all n -blocks repel with a fixed intensity. For sufficiently large n this quantity should be eventually (as $m \rightarrow \infty$) smaller than h^m for any preassigned positive h .

Question 5. As we have mentioned, we only know about conditions which ensure that the limit law for the return time is exponential. It would be interesting to find a (large) class of positive entropy processes for which the distributions of return times are essentially deviated from exponential for bounded away from zero in measure collections of arbitrarily long blocks, i.e., a class of processes with persistent attracting. Can one prove that persistent attracting is, in some reasonable sense, a "typical" property in positive entropy, or that for a fixed measure-preserving transformation with positive entropy, a "typical" generator (partition) leads to persistent attracting?

REFERENCES

- [A-G] Abadi, M. and Galves, A., *Inequalities for the occurrence times of rare events in mixing processes. The state of the art. Inhomogeneous random systems*, Markov Process. Related Fields **7** (2001), 97–112.
- [C-K] Chaumoitre, V. and Kupsa, M., *Asymptotics for return times of rank one systems*, Stochastics and Dynamics **5** (2005), 65–73.
- [C] Coelho, Z., *Asymptotic laws for symbolic dynamical systems*, Topics in symbolic dynamics and applications, London Math. Soc. Lecture Note Ser., vol. 279, Cambridge Univ. Press, 2000, pp. 123–165.
- [D] Downarowicz, T., *Survey of odometers and Toeplitz flows*, Algebraic and Topological Dynamics (Kolyada, Manin, Ward eds), Contemporary Mathematics **385** (2005), 7–38.
- [D-M] Durand, F. and Maass, A., *Limit laws for entrance times for low-complexity Cantor minimal systems*, Nonlinearity **14** (2001), 683–700.
- [G] Gurevič, B. M., *One- and two-sided regularity of stationary random processes*, Dokl. Akad. Nauk SSSR **210** (1973), 763–766.
- [H-L-V] Haydn, N., Lacroix, Y. and Vaienti, S., *Hitting and return times in ergodic dynamical systems*, Ann. Probab. **33** (2005), 2043–2050.
- [H-S-V] Hirata, N., Saussol, B. and Vaienti, S., *Statistics of return times: A general framework and new applications*, Commun. Math. Phys. **206** (1999), 33–55.
- [Kc] Kac, M., *On the notion of recurrence in discrete stochastic processes*, Bull. Amer. Math. Soc. **53** (1947), 1002–1010.
- [Km] Kammerer P., *Das Gesetz der Serie, eine Lehre von den Wiederholungen im Lebens und im Weltgeschehen*, Stuttgart und Berlin, 1919.
- [L] Lacroix, Y., *Possible limit laws for entrance times of an ergodic aperiodic dynamical system*, Israel J. Math. **132** (2002), 253–263.
- [Mi] Von Mises, *Probability, Statistics and Truth*, 2nd ed. rev., New York, Dover, 1981.
- [Mo] Moisset J., *La loi des séries*, JMG Editions, 2000.
- [O] Ornstein, D.S., *Bernoulli shifts with the same entropy are isomorphic*, Adv. Math. **4** (1970), 337–352.
- [O-W1] Ornstein, D.S. and Weiss, B., *Every transformation is bilaterally deterministic*, Israel J. Math. **21** (1975), 154–158.
- [O-W2] Ornstein, D.S. and Weiss, B., *Entropy and recurrence rates for stationary random fields*, IEEE Trans. Inform. Theory **48** (2002), 1694–1697.
- [P] Petersen, K., *Ergodic Theory*, Cambridge Univ. Press, Cambridge, 1983.
- [R1] Rudolph, D.J., *If a two-point extension of a Bernoulli shift has an ergodic square, then it is Bernoulli*, Israel J. Math. **30** (1978), 159–180.
- [R2] Rudolph, D.J., private communication.
- [Sh] Shields, P., *The theory of Bernoulli shifts*, Chicago Lectures in Mathematics, The University of Chicago Press, Chicago, 1973.
- [Si] Sinai, Y., *Weak isomorphism of transformations with an invariant measure*, Sov. Math., Dokl. **3** (1963), 1925–1729; translation from Dokl. Akad. Nauk SSSR 147 (1962), 797–800.
- [Sm] Smorodinsky, M., *On Ornstein's isomorphism theorem for Bernoulli shifts*, Advances in Math. **9** (1972), 1–9.
- [St] Sterzinger, O., *Zur Logik und Naturphilosophie der Wahrscheinlichkeitslehre*, Leipzig, 1919.
- [W] Walters, P., *Ergodic theory—introductory lectures*, Lecture Notes in Mathematics, vol. 458, Springer-Verlag, Berlin, 1975.

INSTITUTE OF MATHEMATICS AND COMPUTER SCIENCE, WROCLAW UNIVERSITY OF TECHNOLOGY, WYBRZEŻE WYSPIAŃSKIEGO 27, 50-370 WROCLAW, POLAND

E-mail address: downar@pwr.wroc.pl

INSTITUT DES SCIENCES DE L'INGÉNIEUR DE TOULON ET DU VAR, AVENUE G. POMPIDOU, B.P. 56, 83162 LA VALETTE DU VAR CEDEX, FRANCE

E-mail address: yves.lacroix@univ-tln.fr